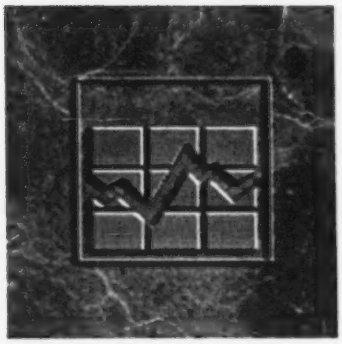
Catalogue no. 62F0026M — No. 001
ISSN 1708-8879
ISBN 978-1-100-18823-2

Research Paper

Household expenditures research paper series

Survey of Household Spending 2009: Data Quality Indicators



Household Survey Methods Division
R. H. Coats Building, Ottawa, K1A 0T6

Telephone: 613-951-7355

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website at www.statcan.gc.ca, e-mail us at infostats@statcan.gc.ca, or telephone us, Monday to Friday from 8:30 a.m. to 4:30 p.m., at the following numbers:

Statistics Canada's National Contact Centre

Toll-free telephone (Canada and United States):

Inquiries line	1-800-263-1136
National telecommunications device for the hearing impaired	1-800-363-7629
Fax line	1-877-287-4369

Local or international calls:

Inquiries line	1-613-951-8116
Fax line	1-613-951-0581

Depository Services Program

Inquiries line	1-800-635-7943
Fax line	1-800-565-7757

To access this product

This product, Catalogue no. 62F0026M, is available free in electronic format. To obtain a single issue, visit our website at www.statcan.gc.ca and browse by "Key resource" > "Publications."

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Statistics Canada
Household Survey Methods Division

Survey of Household Spending 2009: Data Quality Indicators

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2011

All rights reserved. The content of this electronic publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it be done solely for the purposes of private study, research, criticism, review or newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, year of publication, name of product, catalogue number, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, by any means—electronic, mechanical or photocopy—or for any purposes without prior written permission of Licensing Services, Client Services Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

June 2011

Catalogue no. 62F0026M
ISSN 1708-8879
ISBN 978-1-100-18823-2

Frequency: Occasional

Ottawa

Cette publication est disponible en français sur demande.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

This paper was prepared by:

Georgina House
Denis Malo
Marie-Hélène Miville
Sylvain Nadon
Johanne Tremblay

of the Household Survey Methods Division (HSMD).

The authors wish to thank the team responsible for the Survey of Household Spending in the Income Statistics Division (ISD) for their cooperation in the preparation of this report.

Table of Contents

Highlights **6**

Introduction **9**

1. Sampling errors **10**

1.1 Measures of sampling error 10
1.2 Coefficients of variation 11
1.3 Model for deriving an approximation of the CV 12
1.4 Suppression of unreliable data in estimate tables 13

2. Non-response **14**

2.1 Response, non-response and vacancy rates 14
2.2 Non-response by urbanization level 16
2.3 Non-response by income stratum 17
2.4 Adjustment for non-response 17

3. Coverage errors **18**

3.1 Undercoverage and overcoverage: slippage rates 19
3.2 Adjustment at the population and household levels 20

4. Response errors **21**

5. Processing errors **22**

5.1 Proportion of households or individuals requiring imputation, at the national and provincial/territorial levels 24

References **29**

Appendix A Algebraic Notation **30**

Highlights

Sampling errors

- The coefficients of variation (CVs) of the estimates of average total expenditure per household for the provinces vary between 1.4% and 2.3%. The CV at the national level is 1.0%. The CVs are higher in the territories: 5.1% in Yukon, 7.6% in the Northwest Territories, and 6.0% in Nunavut.

- The coefficients of variation (CVs) of the average estimates for the different summary-level expenditure categories are in most cases less than or equal to 2.4% at the national level and generally less than 5.0% at the provincial level. The results for the dwelling and household equipment characteristics are similar. Since the sample size is smaller in Prince Edward Island and the territories, their CVs tend to be higher than the CVs of the other provinces.

Non-response

- The final response rate is 64.5%. The provincial/territorial response rates range from 59.6% in Ontario to 77.9% in Nunavut.

- The non-response rate is 35.5%. Non-response is due to refusals (21.7%), households that could not be contacted (10.8%), and households whose data were considered unusable (3.0%).

- The non-response rate tends to increase with the urbanization level. The non-response rate is 29.5% in rural areas and 38.3% in population centres with a population of 1 million or more.

- According to an analysis of non-response rates in strata of the high-income geographic areas created under the sample design, the rate in high-income strata (40.9%) is higher than the rate observed in other strata (35.0%). The refusal rate for high-income strata (24.7%) is higher than the rate observed for other strata (21.5%).

Coverage errors

- Undercoverage of households is 12.2% at the national level.

- Undercoverage of persons is 16.1% at the national level (excluding the territories).

Processing errors associated with imputation

(i) Expenditure variables

- At least one expenditure variable[1] had to be imputed for 58.9% of households nationally. Contributing strongly to this rate is the section of the questionnaire dealing with expenditures related to residential communications services (telephone, cell phone and Internet access), cable television services, satellite distribution services and security services. For these services, respondent households have been allowed since 2004 to provide only the total expenditure for a package (bundled services), indicating which services are included. Approximately 45% of usable households required imputation for at least one of these six services.

- The overall imputation rate excluding these six services is 29.0% at the national level. At the provincial/territorial level, imputation rates vary from a low of 14.4% for Nunavut to a high of 34.7% for Quebec.

- Imputed values account for 43.8% of the estimate of total expenditure on cable television services and 52.8% of the estimate of total expenditure on Internet access services. Imputation rates for households reporting expenditures on one or the other of these two services are 44.5% and 58.5% respectively. These higher rates are mainly due to the fact that among households that reported paying for a package, a large proportion of packages included these two services.

- About 10% of individuals required imputation of clothing variables, but in the vast majority of such cases, the respondents provided the totals and only the components were imputed.

- About 16% of individuals aged 15 and over required imputation of at least one variable in the Personal Taxes, Security and Money Gifts section.

(ii) Income variables

- Imputation of at least one income variable was required for 7.5% of individuals aged 15 and over

(iii) Categorical variables

- At least one categorical variable had to be imputed for 9.8% of households. Approximately 73% of them had only one variable imputed. At the provincial/territorial level, imputation rates range from a low of 4.3% for Newfoundland and Labrador to a high of 23.9% for the Northwest Territories.

---

1. Excluding the Clothing section and the Personal Taxes, Security and Money Gift section, which are collected at the individual level and not the household level.

Categorical variables that required imputation can be found in the following sections of the questionnaire: Dwelling Characteristics, Facilities Associated with the Dwelling, Tenure, and Tobacco and Miscellaneous, for variables pertaining to purchases through direct sales.

Introduction

The Survey of Household Spending (SHS) is an annual survey that collects data on household income and expenditure using personal interviews. The 2009 SHS sample consists of 16,758 households[2] distributed throughout the ten provinces and three territories. This is a decrease of about 25% in the usual SHS sample size (before 2008), and as a result, the estimates will have higher coefficients of variation. Collection takes place in January, February and March, and income and spending figures are obtained for the period from January 1 to December 31 of the previous year. Following a redesign in 1997, this survey replaces the periodic Family Expenditure Survey and the Household Facilities and Equipment Survey (with modifications to questionnaires and samples).

Like all surveys, the SHS is subject to errors, despite all the precautions taken at the various stages of the survey to control them. While there is no comprehensive measure of the quality of the data generated by a survey, some quality measures produced at the various stages of the survey can provide users with the information they need to interpret the data properly.

This report describes the quality indicators produced for the 2009 Survey of Household Spending. It covers the usual quality indicators that generally help users interpret data, such as coefficients of variation, response and non-response rates, slippage rates and imputation rates.

The quality indicators are grouped by the main types of error in a survey. Section 1 deals with sampling errors, that is, errors due to the fact that the inferences about the population drawn from the survey are based on information collected from a sample of the population, rather than from the entire population. The remaining sections cover non-sampling errors. Non-response and coverage errors are discussed in sections 2 and 3. Response errors and processing errors are dealt with in sections 4 and 5 respectively.

This report focuses on data quality. For a detailed description of the survey's methodology, see reference [1].

2. The initial sample is made up of 20,146 dwellings. Ineligible dwellings are then identified and removed (see section 2.1), leaving the 16,758 households from which data on expenditure and income are collected.

1. Sampling errors

Sampling errors occur when inferences made about a population following a survey are based on information collected from a sample of the population, and not from the entire population. In addition to the sample design and the estimation method used in the Survey of Household Spending, the sample size and the variability of each characteristic are factors that determine sampling error. Characteristics that are rare or are distributed very unevenly in the population will have greater sampling error than characteristics that are observed more frequently or are more homogeneous in the population.

1.1 Measures of sampling error

The standard error is a commonly used measure of sampling error. The standard error is the degree of variation of the estimate due to the fact that one sample was selected rather than another, among all possible samples of the same size under the same sample design. Since the SHS uses a complex sample design and estimation method, the standard error is estimated using a resampling method known as the bootstrap technique. Prior to the 2002 reference year, the jackknife resampling method was used to produce standard error estimates for the SHS. The decision was made to use the bootstrap method for the 2003 and subsequent reference years, mainly because Income Statistics Division (ISD) was planning to publish median expenditure estimates and needed the coefficients of variation of those estimates. The bootstrap method is suitable for variance estimation of non-smooth statistics such as quantiles. For more details on this method, see reference [2].

The coefficient of variation (CV) is also a frequently used measure of the reliability of an estimate. It simply expresses the standard error as a percentage of the estimate. Thus, if an estimate Y is obtained for a certain characteristic and SE is the estimated standard error, then the CV will be (SE/Y) x 100.

In addition, either the standard error or the coefficient of variation may be used to derive another measure of the precision of estimates: the confidence interval. This measure indicates the level of confidence one can have that the true value of an observed characteristic for the population lies within the interval. An interval with a confidence level of 95% corresponds to the estimate obtained from the sample ± 2 standard errors: (Y ± 2 SE).[3] This means that if the sampling were repeated a large number of times, each sample would provide a different interval and 95% of the intervals would contain the true value of the characteristic. Similarly, if the sampling were repeated, the interval Y ± SE would contain the true value in 68% of cases.

3. The confidence interval is calculated directly from the CV in similar fashion: Y ± 2 (CV x Y)/100.

1.2 Coefficients of variation

Estimated coefficients of variation are calculated for estimates of many characteristics collected in the SHS. The CVs of detailed average household expenditure and the CVs of dwelling characteristics and household facilities and equipment are available on request for the national and provincial levels (Income Statistics Division, 1-888-297-7355, income@statcan.gc.ca)

It should be noted that the estimated CVs do not take account of the fact that some of the data were imputed and thus may underestimate the actual CVs. For most variables, the imputation rates are low (see section 5), and the CVs provided are good estimates of the actual CVs. However, to assess the reliability of detailed expenditures with a high imputation rate, we need to consider both the CV and the imputation rate.

Table 1.1 presents an overview of the CVs of estimates of household averages for a few of the summary-level expenditure categories and for income at the provincial/territorial and national levels.

Table 1.1
Coefficients of variation by province, territory and at the national level for the estimation of average household expenditures for several summary level expenditure categories and for the estimation of average income

	Can.	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Yn	N.W.T.	Nt
	%													
Total expenditure	1.0	1.4	1.9	1.8	2.2	1.5	2.0	2.1	1.7	1.7	2.3	5.1	7.6	6.0
Total current consumption	0.8	1.5	1.8	1.6	2.1	1.3	1.4	1.8	1.7	1.8	2.4	4.6	6.8	5.1
Food	0.7	1.4	2.0	1.5	1.8	1.4	1.4	2.2	1.6	2.0	1.9	3.7	4.8	4.3
Shelter	1.0	2.3	2.8	2.3	2.2	2.2	1.8	2.8	2.0	2.4	2.9	6.1	6.2	8.3
Household operation	1.3	2.4	2.9	2.3	4.6	2.3	2.7	3.2	2.5	2.5	2.6	5.5	5.7	9.4
Furnishings	2.3	4.6	6.5	4.7	4.4	4.7	4.2	5.0	5.2	5.6	7.0	10.6	13.3	15.9
Clothing	1.7	2.9	3.9	3.4	3.2	2.9	3.0	3.4	3.3	3.6	5.5	7.7	6.8	6.0
Transportation	1.6	3.7	4.3	4.1	3.6	3.5	2.8	4.3	4.1	4.2	4.1	7.9	12.2	13.1
Health care	1.8	4.4	4.8	3.2	4.1	3.0	4.2	3.9	3.5	3.9	4.3	7.9	9.9	20.6
Personal care	1.4	2.8	3.6	2.5	2.6	2.5	2.6	3.3	3.1	3.3	3.0	6.1	7.6	7.2
Recreation	2.8	4.8	5.9	4.4	12.9	3.6	3.0	3.7	4.5	4.6	15.2	10.1	17.4	7.4
Reading and printed material	2.2	4.9	7.6	4.2	4.4	3.9	4.7	4.4	5.0	4.9	4.1	11.9	14.9	16.1
Education	4.3	14.6	13.8	10.9	11.0	8.0	8.2	10.5	10.4	8.1	9.7	18.3	30.1	60.2
Tobacco, alcoholic beverages	2.3	5.0	7.0	5.0	4.3	3.9	5.2	5.2	5.4	5.5	4.7	9.7	10.5	9.1
Games of chance (net)	5.7	6.8	15.9	8.3	7.7	9.9	6.7	9.9	22.3	28.6	9.0	15.0	18.2	10.8
Miscellaneous expenditures	5.6	6.6	9.2	10.2	20.2	4.8	12.8	6.4	6.7	6.9	10.2	12.1	12.1	28.7
Personal income tax	2.3	2.9	5.0	4.3	3.5	3.8	4.9	5.9	3.9	3.8	4.4	8.1	11.3	10.0
Personal insurance and pension contributions	6.2	2.7	3.6	2.4	2.8	3.2	14.8	2.4	3.5	2.7	2.7	9.9	11.0	8.9
Gifts and contributions	6.9	7.3	14.1	9.7	8.1	20.0	13.0	9.8	9.3	10.0	13.8	14.4	19.0	22.7
Income	0.9	1.2	1.9	1.7	1.5	1.6	2.0	2.1	1.6	1.8	2.2	5.6	8.1	6.2

Table 1.2 presents an overview of the CVs for some dwelling characteristics and household equipment estimates at the provincial/territorial and national levels.

Table 1.2
Coefficients of variation by province, territory and at the national level for some dwelling characteristics and household equipment

	Can.	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Yn	N.W.T.	Nt
	%													
Owner	1.0	2.0	2.8	2.3	1.9	2.6	1.9	2.9	2.3	2.0	2.6	11.6	10.8	18.5
Renter	1.9	7.1	6.6	4.6	6.2	3.6	3.8	5.7	6.4	4.9	4.5	16.2	9.1	4.5
Washing machine	0.8	1.0	1.9	1.7	1.1	1.2	1.7	2.5	1.8	1.5	2.2	5.4	5.1	2.6
Clothes dryer	0.8	1.1	1.8	1.7	1.2	1.4	1.7	2.4	1.6	1.3	2.1	4.9	4.8	2.7
Dishwasher	1.2	3.9	3.3	3.9	3.2	2.6	2.3	3.3	3.4	1.8	2.8	12.8	11.2	16.2
Freezer	1.2	1.8	2.8	2.4	2.0	2.8	2.7	2.9	2.0	2.5	2.9	6.7	5.4	7.9
Microwave oven	0.3	0.6	0.9	0.7	0.8	0.8	0.6	0.8	1.1	0.7	0.9	3.1	1.8	3.4
Cellular phone	0.7	1.9	2.5	2.1	2.0	2.0	1.2	2.1	1.6	1.2	1.3	6.4	7.3	22.7
CD player	0.8	1.5	2.5	1.6	1.5	1.7	1.5	2.3	1.6	1.8	1.7	3.6	4.1	4.5
Cable TV	1.2	2.4	5.8	2.8	3.2	2.4	2.5	3.0	3.8	3.0	2.1	8.5	8.3	12.6
Satellite dish	2.9	5.6	7.4	6.0	3.9	6.7	5.7	6.7	5.9	5.9	9.0	13.1	8.2	15.2
DVD player	0.6	1.3	2.0	1.3	1.4	1.2	1.1	1.3	1.3	1.2	1.1	2.7	3.0	3.5
Home computer	0.7	1.9	2.3	1.6	1.9	1.5	1.2	1.8	1.5	1.3	1.3	5.0	4.3	6.2
Regular tel. connection to a computer (modem)	7.3	25.6	31.5	18.2	17.2	11.5	13.0	16.1	13.6	22.5	24.7	46.6	55.9	55.2
High-speed tel. connection to a computer	2.5	4.7	6.4	4.3	4.0	6.7	4.9	5.2	3.8	6.0	6.4	9.9	16.2	16.4
Cable connection to a computer	2.3	7.9	15.4	7.7	9.6	4.1	4.9	5.9	8.9	5.5	3.9	15.8	12.1	35.8
Use of Internet (home)	0.8	2.1	3.0	1.8	2.2	1.7	1.5	1.9	1.8	1.5	1.3	6.4	4.6	9.4
Owned vehicles (one)	1.8	3.6	5.5	4.2	3.2	3.6	3.7	3.9	6.1	4.3	3.8	9.7	9.8	14.7
Owned vehicles (2 or more)	1.7	4.5	4.4	4.0	2.9	4.0	3.6	3.9	4.3	3.0	3.6	7.7	15.9	26.7

1.3 Model for deriving an approximation of the CV

Estimates for different domains of interest (for example, by income quintile) for the summary-level expenditure categories are available in the publication *Spending Patterns in Canada* (see reference [4]). Estimates for various domains of interest for detailed expenditure categories are available on request from Income Statistics Division. (For more details on tables available on request from Income Statistics Division, see reference [3] or [4].) For operational reasons, it is impossible to produce CVs for all levels of aggregation that may interest users.

1.3.1 Approximation of the CV for domain estimates

However, it is possible to calculate an approximation of the CV by using a relationship between the number of households in the sample that reported expenditures for a given category and the CV for an aggregate level. This relationship, based on the CV's tendency to increase in proportion to a decrease in the square root of the number of households reporting an expenditure, is illustrated below.

Formula for approximating the CV for a domain (a subgroup of the population)

If CV(Y) represents the CV for the estimate of the household average of a certain characteristic for the entire population, then an approximation of the CV for the estimate of that characteristic can be calculated for a domain (which can be regarded as a subgroup of the population, such as a household type, an income quintile or an urbanization level) with the following equation:

$$CV\ (Y_d) = CV\ (Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

where

n : number of households in the sample
P : estimate of the proportion of households reporting a value > 0 for this characteristic in the population
n_d : number of households in the sample in domain d
P_d : estimate of the proportion of households reporting a value > 0 for this characteristic in domain d

In general, approximations for the various domains are calculated using the CV, size n and proportion P at the national level. If an approximation of a CV is desired for a domain that is entirely contained within a single province (for example, a metropolitan area), it is preferable to use these values at the provincial level, since provincial CVs are published for the 2009 SHS (reference [3]). It should be noted that a CV obtained using this approach is only an approximation of the actual value.

1.3.2 Approximation of the CV using the microdata file

Microdata file users can obtain an approximation of the CV of the estimates using another method that will generally provide better results than the method described in the previous section for the CVs of detailed expenditure categories. That approach is described in detail in the documentation provided with the 2009 microdata file. This method of approximation can be used only with the microdata file, since it requires data and weights for each household.

The 1997 SHS data quality report contains the results of the performance evaluation of these two CV approximation methods.

1.4 Suppression of unreliable data in estimate tables

Since the coefficient of variation is an indicator of data reliability, we would like to use it to determine whether the estimates should be published. Estimates whose estimated CV is greater than 33% are not considered sufficiently reliable to be published. However, estimated CVs are not calculated for many of the published

estimates. The suppression rule for expenditure estimates is therefore based on the number of households reporting a value greater than zero.[4]

It can be shown that CVs are usually less than 33% when the number of households reporting an expenditure is greater than 30. Since this is an approximate rule, some estimates may be published even though the CV is greater than 33%, and some estimates will not be published even though the CV is less than 33%. The 1997 SHS data quality report contains the results of the evaluation of the risk of error due to the suppression rule.

2. Non-response

Errors due to non-response occur when potential respondents do not provide the required information or the information they provide is unusable. Cases in which the respondent fails to answer some of the questions are referred to as partial non-response. In those cases, the missing data are imputed. Errors associated with imputation are described in section 5, which deals with processing errors. In this section, non-response includes collection non-response, which is primarily due to inability to contact the household or to refusal by household members to participate in all or part of the survey, as well as data that are collected from households but prove to be unusable.

The main impact of non-response on data quality is that it can cause a bias in the estimates if the characteristics of respondents and non-respondents differ and the difference has an impact on the characteristics studied. Non-response rates are easy to calculate, but they provide only an indication of data quality, since they do not measure the bias associated with the estimates. The magnitude of non-response can be considered an indicator of the risks of bias in the estimates.

2.1 Response, non-response and vacancy rates

Since the units selected in the SHS are dwellings, interviewers must first identify ineligible dwellings, that is, dwellings occupied by persons who are not part of the target population, dwellings that no longer exist (demolished, mobile home moved or dwelling converted to business) and vacant dwellings (unoccupied, seasonal or under construction).

Then, for eligible dwellings, the proportion of households that did not respond to the survey is evaluated. This is called the collection non-response rate. Included are households that refused to participate in the survey and households for which respondents could not be contacted, either because they were absent or because of special circumstances (language problem, illness, death).

Also for eligible dwellings, the rate of unusable data is determined. Unusable data refers to the number of households whose questionnaires were at least partially

4. In practice, we take the estimate of the proportion of households reporting an expenditure and multiply it by the sample size.

completed but were rejected during data processing. When many income or expenditure questions are unanswered, the questionnaire is classified as incomplete and is not used.

Note that all rates provided in this section are unweighted. For the 2009 Survey of Household Spending, the final response rate is 64.5%. Table 2.1-1 shows the final response rate and the sample size (eligible households) broken down by refusals, units not contacted, unusable data and usable data. The information is provided for the national and provincial/territorial levels.

Table 2.1-1
Sample size and response rate by province, territory and at the national level

	Eligible households	No contact	Refusals	Unusables	Usables	Final response rate (at estimation stage)[1]
	number					%
Canada	16,758	1,807	3,644	496	10,811	64.5
Newfoundland and Labrador	1,347	113	207	19	1,008	74.8
Prince Edward Island	789	53	212	9	515	65.3
Nova Scotia	1,515	167	369	51	928	61.3
New Brunswick	1,377	117	290	17	953	69.2
Quebec	2,040	167	505	63	1,305	64.0
Ontario	2,582	371	579	93	1,539	59.6
Manitoba	1,413	182	354	25	852	60.3
Saskatchewan	1,345	114	249	26	956	71.1
Alberta	1,572	196	336	89	951	60.5
British Columbia	1,790	186	406	83	1,115	62.3
Yukon	372	31	65	4	272	73.1
Northwest Territories	376	82	53	11	230	61.2
Nunavut	240	28	19	6	187	77.9

1. Usable/eligible households x 100.

Table 2.1-2 shows the final non-response rate; the collection non-response rate, broken down by refusals and units not contacted; and the rate of households with unusable data owing to incomplete questionnaires. The vacancy rate is also included. The rates are provided for the national and provincial/territorial levels.

Note that the vacancy rates shown in the tables in section 2 include vacant dwellings (unoccupied, seasonal or under construction) and dwellings that no longer exist (demolished, mobile home moved or dwelling converted to business).

Table 2.1-2
Nonresponse and vacancy rates by province, territory and at the national level

	Vacancy rate	Collection non-response rate			Unusable data rate			Final non-response rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
	%							
Canada	**15.0**	**32.5**	**10.8**	**21.7**	**3.0**	**0.7**	**2.3**	**35.5**
Newfoundland and Labrador	15.0	23.8	8.4	15.4	1.4	0.3	1.1	25.2
Prince Edward Island	19.3	33.6	6.7	26.9	1.1	0.4	0.8	34.7
Nova Scotia	18.4	35.4	11.0	24.4	3.4	0.8	2.6	38.7
New Brunswick	16.9	29.6	8.5	21.1	1.2	0.1	1.2	30.8
Quebec	12.2	32.9	8.2	24.8	3.1	0.9	2.2	36.0
Ontario	11.5	36.8	14.4	22.4	3.6	0.6	3.0	40.4
Manitoba	11.5	37.9	12.9	25.1	1.8	0.8	1.0	39.7
Saskatchewan	15.3	27.0	8.5	18.5	1.9	0.1	1.8	28.9
Alberta	14.4	33.8	12.5	21.4	5.7	1.2	4.5	39.5
British Columbia	16.6	33.1	10.4	22.7	4.6	0.8	3.8	37.7
Yukon	17.7	25.8	8.3	17.5	1.1	0.5	0.5	26.9
Northwest Territories	19.8	35.9	21.8	14.1	2.9	1.3	1.6	38.8
Nunavut	20.7	19.6	11.7	7.9	2.5	1.7	0.8	22.1

2.2 Non-response by urbanization level

Non-response varies with urbanization level. The various national rates are shown by urbanization level in Table 2.2.[5]

Table 2.2
Nonresponse and vacancy rates by urbanization level

	Vacancy rate	Collection non-response rate			Unusable data rate			Final non-response rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
Population centre	%							
1,000,000 or more	10.9	34.4	11.6	22.8	3.8	0.8	3.1	38.3
500,000 - 999,999	10.0	36.5	13.1	23.4	3.6	1.0	2.6	40.1
250,000 - 499,999	8.8	38.2	12.4	25.8	4.0	0.6	3.3	42.2
100,000 - 249,999	10.8	33.9	11.4	22.5	2.0	0.4	1.6	35.8
30,000 - 99,999	11.7	32.1	8.5	23.6	2.7	0.7	2.0	34.8
1,000 - 29,999	15.8	28.8	10.9	17.9	2.6	0.6	2.0	31.4
Rural	28.8	27.1	7.6	19.4	2.4	0.5	1.9	29.5
Total	15.0	32.5	10.8	21.7	3.0	0.7	2.3	35.5

5. Tables on non-response rates by urbanization level and province are available on request from Household Survey Methods Division.

2.3 Non-response by income stratum

Since income information is unavailable for non-respondents, it is impossible to compare income-specific non-response rates. However, the LFS sample design, used for the SHS, was constructed in such a way that in census metropolitan areas, there are strata consisting of geographic areas with a high concentration of high-income households. While the number of high-income strata remains relatively small (51 out of a total of 1,060 strata), comparison of non-response rates in this group in relation to the other strata provides relevant information about the potential effect of non-response (see Table 2.3).

Note that in addition to regular strata, the "Other" strata category includes the following four types of strata: strata with a high vacancy rate, high-cost strata, strata with a concentration of immigrants, and strata with a concentration of Aboriginals. Since the portion of the SHS sample allocated to the latter four strata was smaller, the results for them are not broken out in Table 2.3.

Table 2.3
Comparison of non-response and vacancy rates in high-income strata in relation to other strata

	Vacancy rate	Collection non-response rate			Unusable data rate			Final non-response rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
	%							
High-income	8.5	37.9	13.1	24.7	3.1	1.1	2.0	40.9
Other	15.5	32.0	10.6	21.5	2.9	0.6	2.3	35.0
Total	15.0	32.5	10.8	21.7	3.0	0.7	2.3	35.5

2.4 Adjustment for non-response

To compensate for non-response, the weights in the SHS are inflated by the inverse of the weighted response rate within certain predefined groups. Following the overhaul of the LFS sample design, the non-response adjustment groups were redefined. As was the case for previous years, these groups are defined on the basis of the different urbanization levels in each province and of subprovincial geographic areas for Quebec, Ontario and British Columbia. In addition, specific non-response adjustment groups were created for high-income strata. As seen in the previous section, the number of high-income strata increased with the introduction of the new LFS sample design. As a result, it is now possible to form non-response adjustment groups for high-income strata in every province except Prince Edward Island. No such group can be formed in Prince Edward Island because the province has no high-income strata.

The weighted rates differ from the rates presented in this section, since the former take the sampling weight of each household into account. An algebraic description of the non-response adjustment is provided in Appendix A.

The adjustment of weights for non-response takes account of differences in non-response by urbanization level (as illustrated in section 2.2) and geographic area or by groups of high-income strata. It will reduce the bias insofar as the characteristics of respondents and non-respondents are similar for a given urbanization level and geographic area or for a given group of high-income strata.

It should be noted that a non-response adjustment group can be combined with another group if the number of households in the group is too small or the adjustment factor is too high.

3. Coverage errors

The target population was defined in the design of the survey. It is useful to go over this definition, since a good understanding of the target population is necessary in order to properly interpret the survey data. It is important to note that the SHS uses the sampling frame of the Labour Force Survey (LFS).

Target population

> The target population consists of individuals living in private households. It therefore excludes residents of institutions such as prisons, chronic care hospitals and senior citizens' homes, members of religious orders and other groups living communally, members of the Armed Forces living in military compounds, and individuals residing permanently in hotels or rooming houses. Also excluded are foreign countries' official representatives residing in Canada and their families, and individuals residing on Indian reserves or public lands (except for the territories). With these exclusions, the survey covers nearly 98% of the population in the 10 provinces. Coverage is 92% in Yukon, 92% in the Northwest Territories, and 91% in Nunavut.[6]

We did not collect data from persons temporarily living away from their families (for example, students at university) because the information would be obtained from their families if selected in the sample.

Coverage errors result from inadequate representation of the target population by the units in the sampling frame. Some units of the target population may be omitted from the sampling frame, in which case there is undercoverage. Other units that are not in the target population may be included by error, or some units may be included more than once. Those units are responsible for overcoverage.

6. At the household level, coverage is 92% for Yukon, 93% for the Northwest Territories and 92% for Nunavut.

3.1 Undercoverage and overcoverage: slippage rates

In the SHS, the sample is selected using a list of dwellings in each selected cluster. Factors contributing to undercoverage are the omission of dwellings in the creation of the list, new dwellings that are added between the creation of the list and the interviewer's visit (mainly in new developments), and the erroneous classification of vacant dwellings. The inclusion of dwellings that are not within the boundaries of the cluster is a source of overcoverage. Similarly, errors can occur during data collection, due to improper identification of persons as members of the selected household. Those errors also contribute to undercoverage or overcoverage.

In addition, as described in section 2.4, reweighting methods are used to take account of non-response. However, when those adjustments are made, it is impossible to correct the survey weights to ensure that all subgroups within the population are well represented.

Proper representation of the target population is essential to the production of realistic expenditure estimates. The sample must adequately represent the individuals in the target population as well as the distribution of households by size.

There is generally net undercoverage of the number of persons and the number of households in the SHS. That undercoverage is corrected by an adjustment of weights using auxiliary or reference data based on postcensal demographic estimates. The slippage rate (see Appendix A) is a measure of the percentage of difference between the estimates from the auxiliary data and the survey estimates calculated using weights not adjusted with those data.[7] In other words, slippage is the combined effect of undercoverage and imbalance in certain population subgroups created by survey non-response that could not be corrected at the reweighting stage.

Slippage rates by age group for the national and provincial/territorial levels are shown in Table 3.1, while slippage rates by household size, used in adjusting weights, are shown in Table 3.2. A positive rate indicates overcoverage of the number of persons or households in the survey.

7. The subweight, which is the sampling weight adjusted for non-response, is used (see Appendix A).

Table 3.1
Slippage rates for provinces and territories by age group

Age	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia	Canada (excluding the territories)	Yukon	Northwest Territories	Nunavut
	%													
0-6	-10.1	5.2	-0.1	6.7	-4.8	-24.9	-11.2	-15.6	-13.7	-11.7	-14.8	-33.5	-24.0	-11.7
7-17	-9.4	2.7	-13.8	-4.7	-19.2	-16.7	-9.4	-12.6	-11.3	-6.5	-14.4			
18-24	-26.9	-33.4	-33.3	-14.6	-5.8	-23.8	-26.8	-18.7	-26.8	-40.1	-22.5	-31.6	-26.1	-24.3
25-34	-28.0	-4.2	-14.0	-16.7	-13.3	-33.7	-10.4	-7.2	-27.4	-28.6	-25.0			
35-44	-10.6	-8.0	-13.3	-12.3	-14.3	-23.0	-18.5	-11.8	-16.3	-18.0	-18.5			
45-54	-15.4	-17.9	-20.3	-6.9	-24.1	-16.3	-12.4	-6.5	-18.3	-23.2	-18.8			
55-64	0.1	-20.1	-21.2	-9.0	-1.8	-7.8	2.6	-14.3	-20.6	-17.4	-9.0			
65 +	-7.8	17.6	2.0	-1.3	-0.7	-8.4	-1.3	-6.2	-8.3	-9.5	-5.7			
Total	-12.7	-7.3	-14.4	-7.9	-11.2	-19.0	-10.5	-11.0	-18.3	-19.3	-16.1	-32.0	-25.5	-19.4

Table 3.2
Slippage rates for provinces and territories by household size

	Households	One-person households	Two-person households	Households of three or more persons
	%			
Canada	-12.2	-13.6	-4.7	-18.2
Newfoundland and Labrador	-7.0	-4.8	-1.7	-14.5
Prince Edward Island	-5.1	-0.4	-6.6	-6.6
Nova Scotia	-10.3	-6.2	-6.9	-18.1
New Brunswick	-9.2	-20.7	-4.5	-5.7
Quebec	-7.6	-8.5	-1.1	-13.8
Ontario	-13.3	-15.4	-1.5	-21.3
Manitoba	-9.4	-14.6	-4.2	-10.0
Saskatchewan	-6.3	-5.2	-0.4	-13.6
Alberta	-17.5	-17.1	-15.4	-19.6
British Columbia	-17.7	-21.0	-11.6	-20.6
Yukon	-27.1	-18.8	-28.0	-34.2
Northwest Territories	-29.8	-32.2	-42.8	-14.8
Nunavut	-18.3	-33.6	-47.7	3.7

3.2 Adjustment at the population and household levels

To correct the problem of the sample's representativeness, shown in Table 3.1, and to reduce the resulting bias, the survey data are adjusted during weighting using demographic estimates for the age groups defined in this table, for each province and territory. For more details on the adjustment methodology, see

references [1] and [5]. This adjustment reduces the bias but does not eliminate it entirely if the characteristics of the individuals omitted from the survey differ from those of individuals included for a given age group in a province or territory.

It should also be noted that the effectiveness of the adjustment based on demographic estimates depends largely on the quality of those estimates and their accuracy in representing the survey's target population. Demographic estimates are not error-free. They are postcensal estimates based on the population counts from the 2001 Census adjusted for net undercoverage, and they take into account recent statistics on migration, births, deaths, etc. These demographic estimates are adjusted to account for certain exclusions specific to household surveys, such as persons living in institutions. Conceptually, they differ slightly from the SHS target population in that they include persons living in non-institutional collective dwellings, such as members of groups living communally and individuals residing permanently in hotels or rooming houses. However, this difference is considered negligible, since such individuals make up less than 0.4% of the Canadian population.

To remedy the problem of the sample's representativeness in terms of the number of households by size as illustrated in Table 3.2, the survey data are adjusted using auxiliary data. By adjusting the SHS weights to reflect postcensal estimates of the number of households by size, the goal is to compensate for the bias resulting from inadequate representation of households. However, the bias will not necessarily be eliminated if characteristics of households not interviewed (i.e. missed or non-responding households) differ from those of responding households for a given household size. As in the case of demographic estimates of population, the effectiveness of the adjustment will depend on the quality of the auxiliary data on the number of households.

In addition to demographic estimates for age groups by province, two other auxiliary data sets are used during weighting to adjust the survey data and thereby improve their representativeness. The first data set is used to control for the number of children and adults in certain major cities. For the second set, counts for major categories of income from wages and salaries are used in the weight adjustment process to ensure a degree of consistency in income distributions between the SHS and outside sources.

4. Response errors

Response errors represent a lack of accuracy in responses to questions. They can be attributed to various factors, including a questionnaire that requires improvements, misinterpretation of questions by interviewers or respondents, and errors in respondents' statements.

In the SHS, there can be various reasons for errors in respondents' statements. First, there are recall errors that occur when a respondent forgets expenditures made during the period covered by the survey (which corresponds to the calendar year), or when a respondent provides an erroneous value because of the time interval that has elapsed between the time of purchase and the date of the

interview. Recall errors are probably the survey's largest source of response error, since the reference period is long (12 months) and a wide range of information is requested.

To reduce the magnitude of this type of error, the respondent is also encouraged to consult various documents (bills, bank statements, etc.) so as to provide more accurate data. To determine expenditures for small items purchased at regular intervals, interviewers generally suggest that respondents estimate the frequency of the purchases and the price generally paid in order to derive expenditures for a 12-month period.

A second source of error in respondents' reporting is telescopic error, which consists of including in the reference period events that occurred before or after it. In the SHS, the use of the calendar year is considered to provide a good marker for the start of the reference period. Furthermore, since the reference period is a long one, telescopic error has less impact.

Responses by proxy can also contribute to response error. The household member who made an expenditure is generally best able to report it accurately. This is definitely the case with, say, personal purchases. Expenditures reported by an intermediary are more likely to be tainted by response error, and this type of error tends to have a greater effect on certain types of expenditures.

Among other sources of response error, the extent of the respondent's co-operation should not be overlooked. For personal reasons, the respondent may choose not to mention particular expenditures or decide to twist the facts.

In the SHS, another factor is the response burden, owing to the length of the interview and the great variety of items to be reported, as well as the pace of the interview. Interview length varies widely from household to household, depending on household size, income and various other characteristics.

While response errors are a major source of error in a recall interview, they are the most difficult aspect of data quality to measure. Generally, it is necessary to conduct quite costly special studies in an attempt to measure them. Efforts are made to combat response errors by using survey techniques designed to reduce them.

5. Processing errors

Errors can arise in all types of data handling. The main stages of data processing are response coding, data entry, editing, imputation of partial non-response and weighting. In the SHS, different procedures are applied at each stage to minimize processing errors, and the survey estimates are compared with other data sources prior to release. Errors associated with the adjustments made at the weighting stage are described in sections 2 and 3. The other types of processing errors are covered in this section.

Because of the shift to a computer-assisted collection method in 2006, data processing and quality control procedures were altered. Automated edits incorporated into the questionnaire replaced the balance edit checks and edits previously conducted at the regional offices. For the 2009 SHS, interviewers entered responses on a portable computer and conducted an initial edit simultaneously. For example, interval controls, which showed minimums and maximums for certain purchases, were applied if the interviewer entered an unusual amount. Other edits targeted inconsistent responses, such as a household renting its dwelling but paying no rent.

The processing of SHS data also involves imputation for partial non-response. Partial non-response occurs when the respondent refuses to answer or does not know the answer to certain questions. The imputation approach differs depending on whether the data are categorical or continuous. Categorical data take on only specific values (as in yes/no questions or type-of-dwelling questions), while continuous data can take any numerical value (such as for income and expenditure data).

Income and expenditure data are imputed using the nearest neighbour technique. The imputation is done on one group of variables at a time, with the groups formed on the basis of the relationships among the variables. A group generally corresponds to a section of the questionnaire. For each group, the missing values for a recipient (a household that has some missing data for at least one of the variables) are imputed from data on the most similar record among all donors (households that have no missing values for the variables). For each recipient, the closest donor is chosen as the one that minimizes a particular distance function. That function is based on matching variables that are chosen because they are correlated with the variables to be imputed. For example, total household income is chosen as a matching variable for all sections pertaining to expenditures. It is also important to ensure that, after receiving the donor values, the recipient household satisfies certain consistency rules. In general, the imputation is done at the household level, but in some groups (e.g. income and clothing expenditures), the imputation is done at the person level since the original data are collected at that level for those variables.

Since 2001, imputation of income and expenditure data has been done by Statistics Canada's Canadian Census Edit and Imputation System (CANCEIS). This new system is based on methodology that is slightly different from the previous system's methodology. CANCEIS makes better use of categorical variables in the donor selection process. It can also handle the imputation of both continuous and categorical data. CANCEIS was tested prior to its implementation, and the results were similar to those provided by the old system. Since 2003, categorical data, which are found mainly in the dwelling characteristics and facilities sections of the questionnaire, have been imputed with the CANCEIS system. Before that, categorical data were imputed the "hot deck" technique, in which a donor is chosen at random from a group of respondent households with similar characteristics.

The bias caused by imputation of partial non-response is difficult to evaluate. It depends on the differences between respondents and non-respondents and on the imputation method's ability to produce unbiased estimates. However, the

imputation rates provide an indication of the magnitude of partial non-response. They are presented in the next section.

5.1 Proportion of households or individuals requiring imputation, at the national and provincial/territorial levels

An initial indication of the magnitude of partial non-response is the proportion of households requiring imputation and the number of variables imputed per household. The questionnaire can be divided into two major groups of variables: those collected at the household level and those collected at the individual level (such as income and clothing expenditure). For the latter, it is important to note that the respondent may provide only the total income or total clothing expenditures if he/she is unable to provide the breakdowns by source of income or type of expenditure. This increases the level of imputation for the components of income and clothing expenditure, but it does not affect total income, total clothing expenditure or total expenditure.

The percentage of households requiring imputation for household expenditure (excluding clothing expenditures and expenditures in the section on Personal Taxes, Security and Money Gifts) is presented in the next subsection. The subsection after that presents the percentage of persons requiring imputation for a clothing expenditure variable, the percentage of persons requiring imputation for an income variable and the percentage of persons requiring imputation for a variable in the section on Personal Taxes, Security and Money Gifts. The last subsection presents the results for the percentage of households requiring imputation of at least one categorical variable. After data imputation by the system, some corrections might be needed in both imputed and non-imputed variables to ensure data consistency. In reality, such changes constitute only a very small percentage. The results are provided for the national and provincial/territorial levels. That gives an indication of which provinces or territories are most affected by imputation.

5.1.1 Household expenditure imputation for the provinces and territories

Table 5.1-1 shows the percentage of usable households requiring imputation of at least one expenditure variable. Usable households are all households living in eligible dwellings, excluding households that could not be contacted, refused to participate in the survey or provided incomplete data (see definitions in section 2.1). The table is broken down by the number of imputed variables (out of 259) for a household.

Note that regular mortgage payments and mortgage insurance premiums are included under shelter costs and thus under total expenditure. In 2002, those two variables were added to the calculation of imputation rates shown in Table 5.1-1. The impact of that change is a higher overall imputation rate.

In 2004, a change was made in the part of the questionnaire concerning expenditures on residential communication services (telephone, cell phone and Internet access), cable television services, satellite distribution services and security systems. Because of the growing use of packages (bundled services), a household may be billed for combined services and may therefore be unable to

provide expenditures for individual services. In such cases, respondent households may provide only the total expenditure for those services, while indicating which services are included in the package. Expenditures for individual services are then imputed in two stages. First, we impute households for which only a few services are missing, followed by households for which only the total expenditure for the package is available. For the latter households, the imputed expenditures for services (those included in the package) are adjusted proportionally so that their sum corresponds to the total expenditure on the package reported by the respondent household. Since this change has had a major impact on the overall imputation rate for expenditures, the imputation rates in Table 5.1-1 are shown separately with and without the costs of residential communications services, rental of cable television services, rental of satellite distribution services and rental of security services. Also, since this change has had an impact on the level of imputation of expenditures for the six services, Table 5.1-2 is provided, showing the imputation rate and a measure of the impact of imputation for each service.

Table 5.1-1
Households requiring expenditure imputation by province, territory and at the national level

	Households requiring imputation of **expenditure variables**[1] (excluding clothing expenditures and expenditures in the section on Personal Taxes, Security and Money Gifts)							
	Excluding expenditures related to residential communications services, cable and satellite television services and security systems				Including expenditures related to residential communications services, cable and satellite television services and security systems			
	Number of variables imputed (out of 259)			Total	Number of variables imputed (out of 265)			Total
	1	2	3 or more		1	2	3 or more	
	%							
Canada	14.9	6.4	7.6	29.0	7.5	17.7	33.7	58.9
Newfoundland and Labrador	15.1	4.1	5.0	24.1	6.6	17.7	30.7	55.0
Prince Edward Island	17.3	7.2	4.9	29.3	7.4	17.7	42.7	67.8
Nova Scotia	15.3	7.3	7.2	29.8	5.9	17.9	43.5	67.3
New Brunswick	18.4	6.8	7.1	32.3	9.5	16.7	32.7	59.0
Quebec	18.8	6.2	9.7	34.7	9.1	13.9	39.8	62.8
Ontario	12.7	7.8	8.3	28.7	7.7	14.2	34.3	56.2
Manitoba	14.6	7.7	8.2	30.5	8.1	16.3	30.3	54.7
Saskatchewan	15.0	6.6	7.0	28.6	5.6	20.3	39.6	65.6
Alberta	11.7	5.5	8.6	25.8	6.7	21.2	31.1	59.1
British Columbia	13.5	7.0	7.9	28.4	6.3	25.5	30.0	61.7
Yukon	17.6	3.3	5.1	26.1	12.1	21.7	14.0	47.8
Northwest Territories	10.4	3.0	14.3	27.8	7.8	13.9	17.8	39.6
Nunavut	8.6	3.7	2.1	14.4	8.6	4.3	3.2	16.0

1. Includes regular mortgage payments and mortgage insurance premiums.

Since there is a higher level of imputation for expenditures related to residential communications services, cable television services, satellite distribution services and security systems, it is important to measure the effect that imputation has on the estimates of totals for those six variables. This measure, along with the imputation rate, can be used to track the amount of imputation of those variables over time. Because of the growing popularity of packages (bundled services), the imputation level should increase over time. To measure the impact of imputation,

the weighted total of the imputed data is divided by the total estimate (sum of weighted values). This measure represents the proportion of the total value of the estimate that is obtained from imputed data.

Table 5.1-2
Impact of imputation on communications services, cable television services, satellite distribution services and security systems at the national level

	Impact of imputation	Reporting households requiring imputation of expenditure for residential communications services, cable television services, satellite distribution services or security systems
	%	
Telephone services	34.7	40.8
Cell phone, pager and handheld text messaging services	11.5	13.2
Cable television services	43.8	44.5
Satellite distribution services	14.4	11.8
Internet access services	52.8	58.5
Security systems	4.5	7.6

5.1.2 Individual expenditure and income imputation for the provinces and territories

Since some respondents provide only totals for the clothing expenditure and income variables, a two-step procedure is used to impute the variables (at the individual level). Individuals who require imputation of only certain components are imputed first, followed by those for whom only totals are available but imputation of all components is required. (See reference [1] for a more detailed description of this process.)

The percentage of usable individuals (members of usable households) requiring imputation of at least one income variable is presented for the provinces and territories in Table 5.2. The table shows the percentage of persons who had exactly one variable imputed, the percentage who had two or more variables (but not all) imputed and the percentage of persons for whom only total income was available (and hence had to have all their components imputed). The total percentage of persons requiring some form of income imputation is also provided. The second-last column of Table 5.2 indicates the total percentage of persons requiring some form of imputation for clothing expenditure variables. The last column shows the total percentage of persons requiring some form of imputation for the Personal Taxes, Security and Money Gifts section of the questionnaire.

Note that only household members aged 15 or over on December 31 of the reference year are required to answer the questions relating to personal income, personal taxes, security and money gifts. Thus, since the 2003 reference year, the

percentage of persons requiring some form of imputation for income variables as well as for the Personal Taxes, Security and Money Gifts section has been calculated using only persons aged 15 or over, rather than all persons as in previous years. This modification resulted in a slightly higher imputation rate for those variables. As in previous years, the percentage of persons requiring imputation for clothing expenditure variables is based on all persons, since those expenditure questions are asked for each household member.

Table 5.2
Persons requiring income imputation, persons requiring clothing expenditure imputation and persons requiring imputation for variables in personal taxes, security and money gifts section by province, territory and at the national level

	Percentage of persons requiring imputation of income variables				Percentage of persons requiring imputation of at least one of the 15 clothing variables	Percentage of persons requiring imputation of at least one of the 15 variables in the Personal Taxes, Security and Money Gifts section.
	1 variable imputed	2 or more variables imputed (not all)	All variables imputed (total income known)	**Total** (any form of imputation)		
	%					
Canada	4.6	2.1	0.6	7.5	10.3	16.0
N.L.	2.6	1.0	0.7	4.4	12.1	9.6
P.E.I.	4.5	4.1	1.0	10.0	12.8	18.0
N.S.	4.0	2.9	0.9	7.9	17.2	19.7
N.B.	3.8	1.3	0.5	5.9	7.7	17.9
Que.	6.2	2.2	0.2	8.6	9.6	16.1
Ont.	5.2	2.3	0.3	7.8	6.1	17.3
Man.	7.0	2.5	0.3	9.9	6.5	16.4
Sask.	4.6	2.5	0.8	8.0	7.4	15.7
Alta.	4.3	2.0	0.6	7.1	7.0	16.8
B.C.	5.1	2.5	1.0	8.7	18.4	18.7
Yukon	4.7	0.6	0.4	5.7	9.6	11.2
Northwest Territories	1.7	0.8	0.0	2.7	21.3	8.5
Nunavut	1.8	0.9	0.2	2.9	5.8	5.4

5.1.3 Imputation of categorical variables for the provinces and territories

Table 5.3 shows the percentage of usable households requiring imputation of at least one categorical variable. The table is broken down by the number of imputed variables (out of 42) for a household. Categorical variables that required imputation can be found in the following sections of the questionnaire: Dwelling Characteristics (with the exception of the dwelling type variable); Facilities Associated with the Dwelling; Tenure (with the exception of variables related to a tenure change during the reference year); Tobacco and Miscellaneous for variables pertaining to purchases through direct sales (yes/no questions). Note that other categorical variables from the questionnaire, such as the household composition variables or questionnaire skips, are edited and validated by subject-

matter experts from Income Statistics Division. Consequently, the latter variables are not imputed using the nearest neighbour technique.

Table 5.3
Households requiring imputation of categorical variables by province, territory and at the national level

	Number of variables imputed (out of 42)			
	1	2	3 or more	Total
	%			
Canada	7.2	1.6	1.0	9.8
Newfoundland and Labrador	4.1	0.1	0.1	4.3
Prince Edward Island	10.5	0.8	1.7	13.0
Nova Scotia	6.0	1.4	0.4	7.9
New Brunswick	6.2	1.6	0.0	7.8
Quebec	6.8	1.1	0.8	8.7
Ontario	6.9	2.2	2.5	11.6
Manitoba	10.2	3.4	0.9	14.6
Saskatchewan	6.1	2.0	0.8	8.9
Alberta	6.9	1.3	0.6	8.8
British Columbia	8.4	1.7	1.2	11.3
Yukon	8.8	0.7	1.5	11.0
Northwest Territories	16.1	5.2	2.6	23.9
Nunavut	4.3	0.0	0.5	4.8

References

[1] Tremblay, J. and Arsenault, S. 2001. *Methodology of the Survey of Household Spending.* Catalogue no. 62F0026MIE2001003. Ottawa. Household Survey Methods Division, Statistics Canada.

[2] Rao, J.N.K., Wu, C.F.J., Yue, K. 1992. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada, Catalogue no. 12-001); 18(2):209-217.

[3] Statistics Canada, Income Statistics Division. 2009. *User Guide for the Survey of Household Spending.* Catalogue no. 62F0026MIE - N° 001. Ottawa.

[4] Statistics Canada, Income Statistics Division. 2009. *Spending Patterns in Canada*. Catalogue no. 62-202. Ottawa.

[5] Lessard, S. 2005. *Revision of the Calibration Strategy for the Survey of Household Spending.* Ottawa. Household Survey Methods Division, Statistics Canada.

Appendix A

Algebraic Notation

1. Non-response adjustment

The subweight (i.e. the design weight adjusted for non-response) for a household k, denoted as w_k^{NR}, is

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{rate_g} \quad with \quad rate_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

where

$s_{g,r}$ is the set of respondent households in non-response group g,

$s_{g,nr}$ is the set of non-respondent households (refusals, units not contacted, unusable data) in non-response group g, and

π_k^{-1} is the design weight attributed to household k.

2. Calculation of the slippage rate

The slippage rate for a control group c, denoted as $rate_c$, is

$$rate_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR} \right) - t_c}{t_c}$$

where

$s_{c,r}$ is the set of respondent households in control group c,

w_k^{NR} is the subweight of household k, and

t_c is the total of the auxiliary data for control group c.